

Calidad del Agua

Micaela Jhoselin Saenz Molina
msaenzm@fcpn.edu.bo

Encargado:

Moises Martin Silva Choque, M.Sc²
msilva@fcpn.edu.bo

Universidad de San Andrés, Facultad de Ciencias Puras y Naturales, Programa de INF-354 Inteligencia Artificial

Resumen

El agua contaminada y el saneamiento deficiente están relacionados con la transmisión de enfermedades como el cólera, la diarrea, la disentería, la hepatitis A, la fiebre tifoidea y la poliomielitis. Los servicios de agua y saneamiento inexistentes, inadecuados o mal administrados exponen a las personas a riesgos de salud prevenibles. Este es particularmente el caso en los establecimientos de salud donde tanto los pacientes como el personal corren un riesgo adicional de infección y enfermedad cuando faltan servicios de agua, saneamiento e higiene. A nivel mundial, el 15% de los pacientes desarrollan una infección durante una estadía en el hospital, con una proporción mucho mayor en los países de bajos ingresos.

Entonces, me inspiré un poco en esto para usar este conjunto de datos de calidad del agua para comprender qué constituye agua potable segura y aplicarle el aprendizaje automático para distinguir entre agua potable y no potable.

Palabras Clave: Agua, Contaminada, Modelos, Inteligencia Artificial

1. INTRODUCCIÓN

La calidad del agua son las características químicas, físicas y biológicas del agua basadas en estándares de uso. Se utilizan con mayor frecuencia en relación con un conjunto de estándares, cuyo cumplimiento, por regla general, se logra mediante el tratamiento del agua, se puede evaluar. Los estándares más comunes utilizados para monitorear y evaluar la calidad del agua reflejan la salud de los ecosistemas, la seguridad del contacto humano y la salud del agua potable. La calidad del agua tiene un impacto significativo en el suministro de agua y, a menudo, determina las opciones de suministro. Los parámetros de calidad del agua están determinados por el uso previsto. El trabajo de calidad del agua tiende a centrarse en el agua que se trata para agua potable, uso industrial/doméstico o restauración (medio ambiente/ecosistema, típicamente salud humana/vida acuática).

1.1 Objetivos

Entre los objetivos tendremos el lograr entender de forma correcta el dataset además de producir un modelo predatorio aceptado, bajo una precisión aceptable. Esto se lograra mediante el entendimiento de los datos provistos y como están distribuidos

1.2 Contexto

El agua salubre y fácilmente accesible es importante para la salud pública, tanto si se utiliza para beber, para uso doméstico, para producir alimentos o para fines recreativos. La mejora del abastecimiento de agua, del saneamiento y de la gestión de los recursos hídricos puede impulsar el crecimiento económico de los países y contribuir en gran medida a la reducción de la pobreza.

1.2.1. Importancia de la Investigación

El agua segura y suficiente facilita la práctica de la higiene, que es una medida clave para prevenir no solo enfermedades diarreicas, sino también infecciones respiratorias agudas y numerosas enfermedades

tropicales desatendidas.

En el mundo hay al menos 2000 millones de personas que utilizan una fuente de agua para consumo humano contaminada con heces. La contaminación microbiana del agua para estos fines como resultado de la contaminación con heces supone el mayor riesgo en cuanto a salubridad y transmisión de enfermedades como la diarrea, el cólera, la disentería, la fiebre tifoidea y la poliomielitis.

Más de 2000 millones de personas viven en países con escasez de agua, situación que probablemente empeorará en algunas regiones como resultado del cambio climático y el crecimiento de la población.

En 2019, en los países menos adelantados, solo el 50% de los establecimientos de salud tenían servicios básicos relacionados con el agua; el 37%, servicios básicos de saneamiento, y el 30%, servicios básicos de gestión de residuos.

El agua para consumo humano microbiológicamente contaminada puede transmitir todas esas enfermedades y, según se calcula, causa 485 000 muertes por diarrea cada año. Aunque los riesgos químicos más importantes para este tipo de agua provienen del arsénico, el fluoruro o el nitrato, nuevos contaminantes, como productos farmacéuticos, pesticidas, sustancias perfluoroalquiladas y polifluoroalquiladas (PFAS) y los microplásticos son motivo de preocupación pública.

En 2020, el 74% de la población mundial (5800 millones de personas) utilizaba un servicio de suministro de agua para consumo humano gestionado de forma segura —es decir, ubicado en el lugar de uso, disponible cuando se necesita y no contaminado.

1.2.2. Descripción del Dataset

El dataset viene de analizar, las métricas de calidad de agua de 3276 cuerpos de agua, este consta de 10 columnas que analizan los siguientes valores.

ph: pH de 1. agua (0 a 14).

Dureza: Capacidad del agua para precipitar jabón en mg/L.

Sólidos: Sólidos disueltos totales en ppm.

Cloraminas: Cantidad de Cloraminas en ppm.

Sulfato: Cantidad de Sulfatos disueltos en mg/L.

Conductividad: Conductividad eléctrica del agua en $\mu\text{S}/\text{cm}$.

Organic_carbon: Cantidad de carbono orgánico en ppm.

Trihalometanos: Cantidad de Trihalometanos en $\mu\text{g}/\text{L}$.

Turbidez: Medida de la propiedad de emisión de luz del agua en NTU.

Potabilidad: Indica si el agua es segura para el consumo humano. Potable - 1 y No potable – 0

Cada uno de estos valores, son clasificados bajo las métricas anteriormente explicadas, cada una de las cuales funciona de la siguiente manera, bajo los siguientes conceptos:

1. valor pH:

El PH es un parámetro importante para evaluar el equilibrio ácido-base del agua. También es el indicador de la condición ácida o alcalina del estado del agua. La OMS ha recomendado un límite máximo permisible de pH de 6,5 a 8,5. Los rangos de investigación actuales fueron de 6,52 a 6,83, que se encuentran en el rango de los estándares de la OMS.

2. Dureza:

La dureza es causada principalmente por sales de calcio y magnesio. Estas sales se disuelven a partir de depósitos geológicos a través de los cuales viaja el agua. El tiempo que el agua está en contacto con el material que produce dureza ayuda a determinar cuánta dureza hay en el agua cruda. La dureza se definió originalmente como la capacidad del agua para precipitar el jabón causado por el Calcio y el Magnesio.

3. Sólidos (Sólidos disueltos totales - TDS):

El agua tiene la capacidad de disolver una amplia gama de minerales o sales inorgánicos y algunos orgánicos, como potasio, calcio, sodio, bicarbonatos, cloruros, magnesio, sulfatos, etc. Estos minerales producen un sabor no deseado y un color diluido en la apariencia del agua. Este es el parámetro importante para el uso del agua. El agua con alto valor de TDS indica que el agua está altamente mineralizada. El límite deseable para TDS es 500 mg/l y el límite máximo es 1000 mg/l que se prescribe para beber.

4. Cloraminas:

El cloro y la cloramina son los principales desinfectantes utilizados en los sistemas públicos de agua. Las cloraminas se forman más comúnmente cuando se agrega amoníaco al cloro para tratar el agua potable. Los niveles de cloro de hasta 4 miligramos por litro (mg/L o 4 partes por millón (ppm)) se consideran seguros en el agua potable.

5. Sulfato:

Los sulfatos son sustancias naturales que se encuentran en los minerales, el suelo y las rocas. Están presentes en el aire ambiente, las aguas subterráneas, las plantas y los alimentos. El principal uso comercial del sulfato es en la industria química. La concentración de sulfato en el agua de mar es de aproximadamente 2700 miligramos por litro (mg/L). Varía de 3 a 30 mg/L en la mayoría de los suministros de agua dulce, aunque se encuentran concentraciones mucho más altas (1000 mg/L) en algunas ubicaciones geográficas.

6. Conductividad:

El agua pura no es un buen conductor de la corriente eléctrica sino un buen aislante. El aumento de la concentración de iones mejora la conductividad eléctrica del agua. Generalmente, la cantidad de sólidos disueltos en el agua determina la conductividad eléctrica. La conductividad eléctrica (CE) en realidad mide el proceso iónico de una solución que le permite transmitir corriente. De acuerdo con los estándares de la OMS, el valor de EC no debe exceder los 400 $\mu\text{S}/\text{cm}$.

7. Carbono orgánico:

El carbono orgánico total (TOC) en las fuentes de agua proviene de la materia orgánica natural en descomposición (NOM), así como de fuentes sintéticas. TOC es una medida de la cantidad total de carbono en compuestos orgánicos en agua pura. Según US EPA $< 2 \text{ mg/L}$ como TOC en agua tratada/bebible, y $< 4 \text{ mg/Lt}$ en agua de fuente que se usa para tratamiento.

8. Trihalometanos:

Los THM son sustancias químicas que se pueden encontrar en el agua tratada con cloro. La concentración de THM en el agua potable varía según el nivel de materia orgánica en el agua, la cantidad de cloro requerida para tratar el agua y la temperatura del agua que se está tratando. Los niveles de THM de hasta 80 ppm se consideran seguros en el agua potable.

9. Turbidez:

La turbidez del agua depende de la cantidad de materia sólida presente en estado de suspensión. Es una medida de las propiedades emisoras de luz del agua y la prueba se utiliza para indicar la calidad de la descarga de desechos con respecto a la materia coloidal. El valor medio de turbidez obtenido para Wondo Genet Campus (0,98 NTU) es inferior al valor recomendado por la OMS de 5,00 NTU.

10. Potabilidad:

Indica si el agua es segura para el consumo humano donde 1 significa Potable y 0 significa No potable.

2. MATERIALES Y MÉTODOS/METODOLOGÍA

Para realizar la selección, del modelo así también como para lograr la satisfacción del objetivo debemos de conocer bien

los datos que tenemos, así como también entender su importancia. Iniciaremos con una de las características mas destacables y mejor conocida, como es la potabilidad.

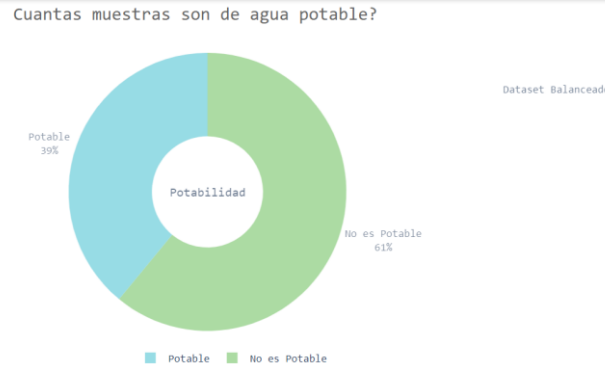


Figura 1 Potabilidad

Procederemos a ver las otras características, basados en el primer análisis del agua que es potable.

Dureza del agua: La definición simple de dureza del agua es la cantidad de calcio y magnesio disueltos en el agua. El agua dura tiene un alto contenido de minerales disueltos, principalmente calcio y magnesio. Es posible que haya sentido los efectos del agua dura, literalmente, la última vez que se lavó las manos. Dependiendo de la dureza de su agua, después de usar jabón para lavarse puede haber sentido que quedaba una película de residuos en sus manos. En agua dura, el jabón reacciona con el calcio (que es relativamente alto en agua dura) para formar "escoria de jabón". Cuando se usa agua dura, se necesita más jabón o detergente para limpiar las cosas, ya sean las manos, el cabello o la ropa.

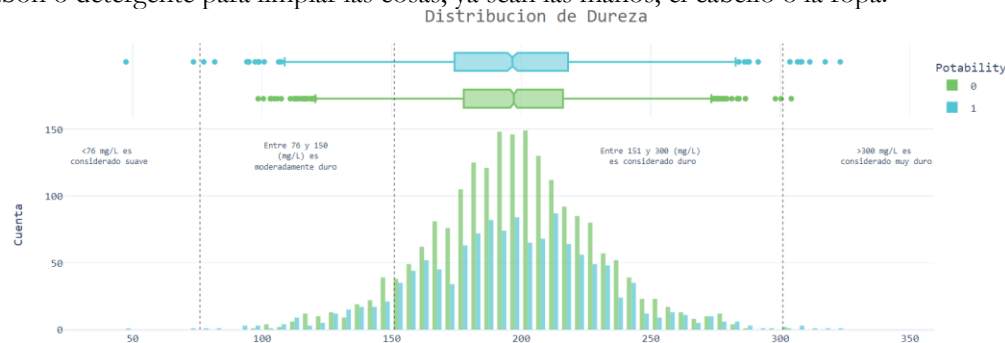


Figura 2 Distribución de dureza

Por otro lado, tenemos el pH del agua es una medida del equilibrio ácido-base y, en la mayoría de las aguas naturales, está controlado por el sistema de equilibrio dióxido de carbono-bicarbonato-carbonato. Por lo tanto, una mayor concentración de dióxido de carbono disminuirá el pH, mientras que una disminución hará que aumente. La temperatura también afectará los equilibrios y el pH. En agua pura, se produce una disminución del pH de aproximadamente 0,45 cuando la temperatura aumenta 25 °C. En aguas con capacidad amortiguadora impartida por iones bicarbonato, carbonato e hidroxilo, este efecto de la temperatura se modifica (APHA, 1989). El pH de la mayor parte del agua potable se encuentra dentro del rango de 6,5 a 8,5. Las aguas naturales pueden tener un pH más bajo, como resultado, por ejemplo, de la lluvia ácida o un pH más alto en áreas calizas.

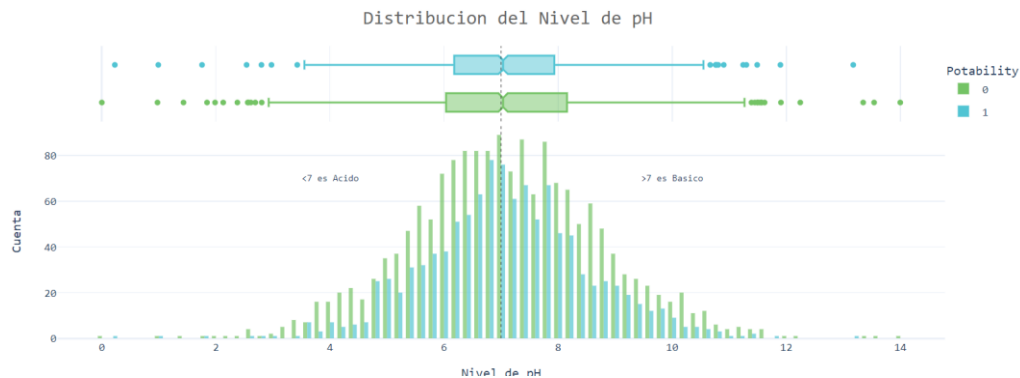


Figura 3 Nivel de pH

Así también, tenemos TDS que significa concentración de partículas disueltas o sólidos en agua. TDS se compone de sales inorgánicas como calcio, magnesio, cloruros, sulfatos, bicarbonatos, etc., junto con muchos más compuestos inorgánicos que se disuelven fácilmente en agua.

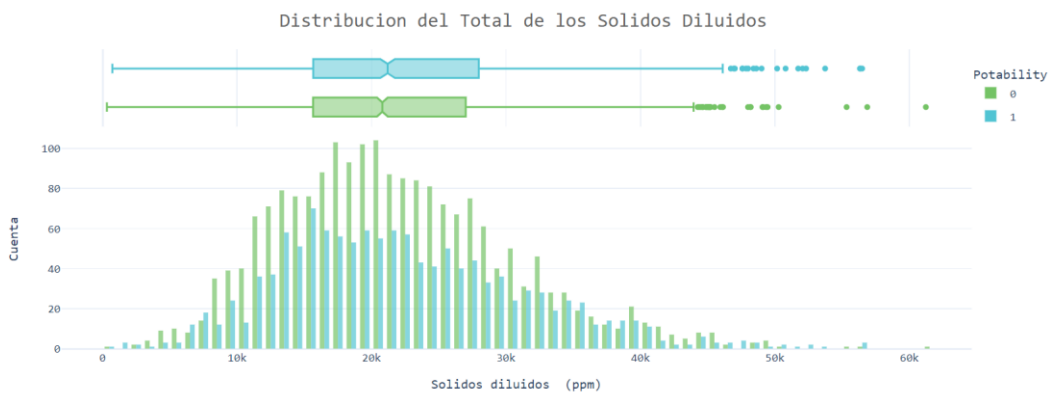


Figura 4 Distribución de TDS

Continuando, analizaremos la distribución de las cloraminas (también conocidas como desinfección secundaria) son desinfectantes que se utilizan para tratar el agua potable y:

- Se forman más comúnmente cuando se agrega amoníaco al cloro para tratar el agua potable.
- Proporcione una desinfección más duradera a medida que el agua se mueve a través de las tuberías hacia los consumidores.

Las empresas de servicios de agua han utilizado cloraminas desde la década de 1930.

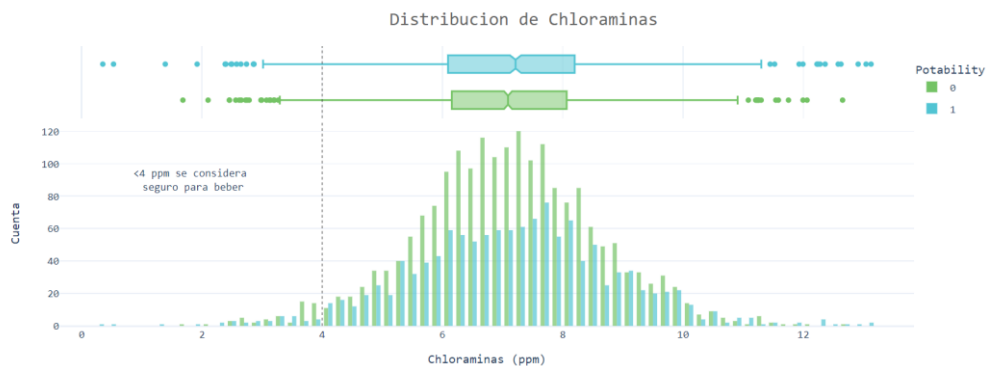


Figura 5 Distribución de Cloraminas

De la misma forma, visualizaremos el sulfato (SO_4) se puede encontrar en casi todas las aguas naturales. El origen de la mayoría de los compuestos de sulfato es la oxidación de minerales de sulfito, la presencia de lutitas o los desechos

industriales. El sulfato es uno de los principales componentes disueltos de la lluvia. Las altas concentraciones de sulfato en el agua que bebemos pueden tener un efecto laxante cuando se combinan con calcio y magnesio, los dos componentes más comunes de la dureza.

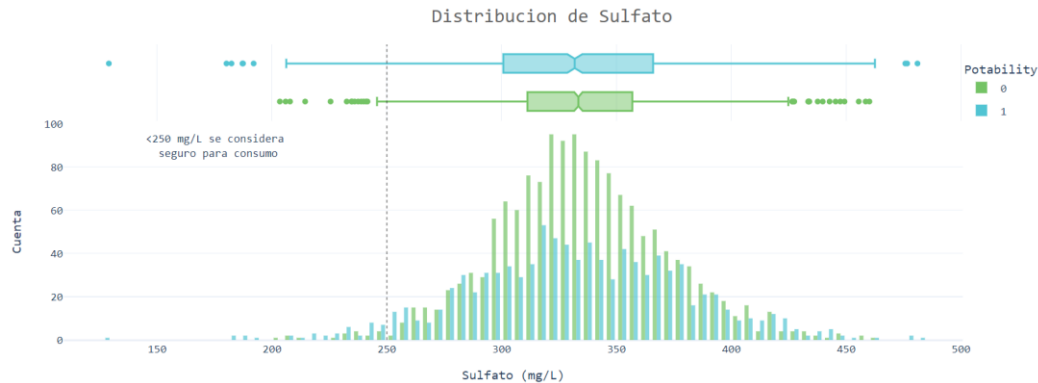


Figura 6 Distribución de Sulfato

Asimismo, revisaremos la conductividad es una medida de la capacidad del agua para pasar una corriente eléctrica. Debido a que las sales disueltas y otras sustancias químicas inorgánicas conducen la corriente eléctrica, la conductividad aumenta a medida que aumenta la salinidad. Los compuestos orgánicos como el aceite no conducen muy bien la corriente eléctrica y, por lo tanto, tienen una baja conductividad en el agua. La conductividad también se ve afectada por la temperatura: cuanto más caliente esté el agua, mayor será la conductividad.

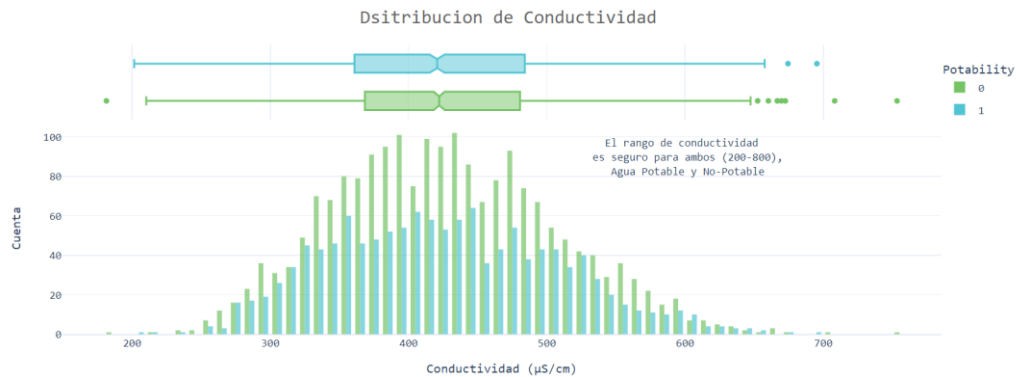


Figura 7 Distribución de Conductividad

De esta manera, los contaminantes orgánicos (sustancias orgánicas naturales, insecticidas, herbicidas y otros productos químicos agrícolas) ingresan a las vías fluviales en la escorrentía de las lluvias. Las aguas residuales domésticas e industriales también aportan contaminantes orgánicos en diversas cantidades. Como resultado de derrames o fugas accidentales, los desechos orgánicos industriales pueden ingresar a los arroyos. Es posible que algunos de los contaminantes no se eliminen por completo mediante procesos de tratamiento; por lo tanto, podrían convertirse en un problema para las fuentes de agua potable. Es importante conocer el contenido orgánico en una vía fluvial.

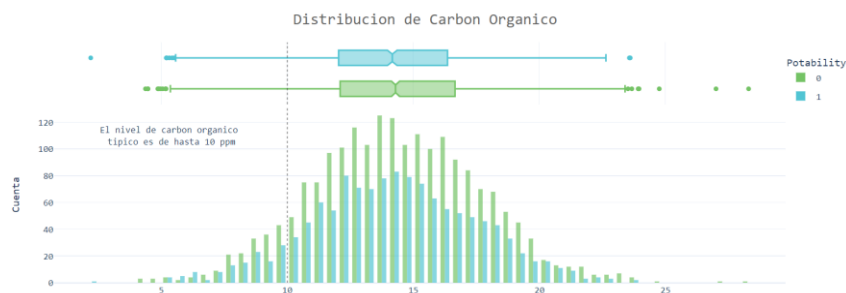


Figura 8 Distribución de Contaminantes Orgánicos

Como penúltima característica, los trihalometanos (THM) son el resultado de una reacción entre el cloro utilizado para desinfectar el agua del grifo y la materia orgánica natural del agua. En niveles elevados, los THM se han asociado con efectos negativos para la salud, como cáncer y resultados reproductivos adversos.

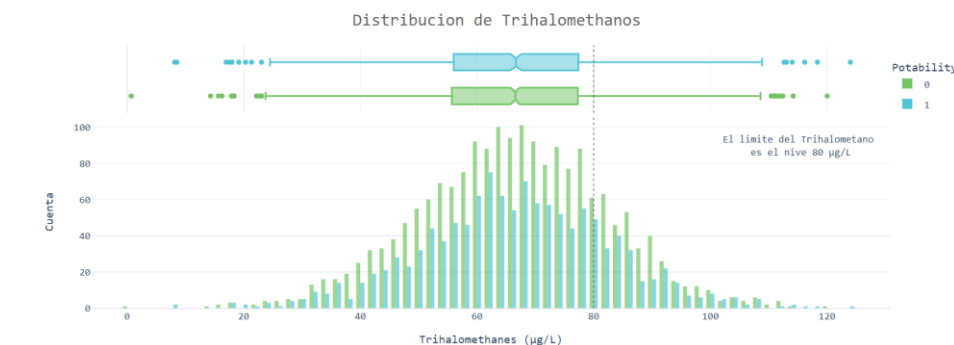


Figura 9 Distribución de THM

Finalmente, como ultimo análisis de los datos tenemos la turbidez es la medida de la claridad relativa de un líquido. Es una característica óptica del agua y es una medida de la cantidad de luz que es dispersada por el material en el agua cuando se hace brillar una luz a través de la muestra de agua. Cuanto mayor sea la intensidad de la luz dispersada, mayor será la turbidez. El material que hace que el agua se vuelva turbia incluye arcilla, limo, materia inorgánica y orgánica muy pequeña, algas, compuestos orgánicos coloreados disueltos y plancton y otros organismos microscópicos.

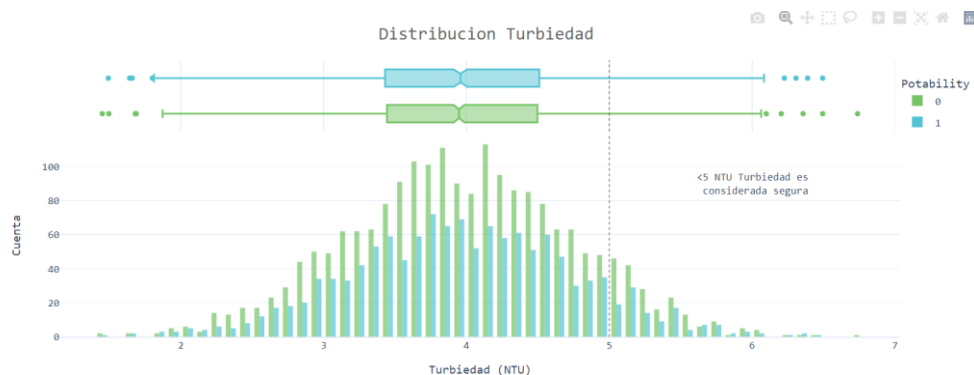


Figura 10 Distribución de turbiedad

Debemos de mencionar, que todos los anteriores gráficos son interactivos y este proyecto analizo a los datos unidos de otras formas además podemos continuar observando el resto de esto en el desarrollo de GitHub de este proyecto. Después de proceder con una limpieza general de la base de datos, así como también una estandarización general de los datos, hemos usado el método Stop Checking para encontrar el mejor modelo. Hemos seleccionado el modelo de random forest y aunque la precisión es 0.66 y aunque es difícil para este modelo diferenciar bien las características de esta base de datos.

3. RESULTADOS Y DISCUSIÓN

Acerca del análisis con gráficos, y el uso de random forest este proceso ciertamente fue tardado incluso si la

precisión realmente no fue de las mejores, es por eso que he llegado al resumen los siguientes puntos importantes de la primera ronda de análisis.

Los niveles de TDS parecen contener alguna descripción ya que sus valores son en promedio 40 veces más que el límite superior para agua potable segura.

Los datos contienen casi el mismo número de muestras de agua con niveles de pH ácidos y básicos.

El 92% de los datos se consideró duro.

Solo el 2% de las muestras de agua eran seguras en términos de niveles de cloraminas.

Solo el 1,8 % de las muestras de agua eran seguras en cuanto a los niveles de sulfato.

El 90,6 % de las muestras de agua tenían niveles de carbono más altos que los niveles típicos de carbono en el agua potable (10 ppm).

El 76,6% de las muestras de agua eran seguras para beber en términos de niveles de trihalometano en el agua.

El 90,4% de las muestras de agua eran seguras para beber en términos de turbidez de las muestras de agua.

Los coeficientes de correlación entre las características fueron muy bajos.

Random Forest y XGBoost funcionaron mejor para entrenar el modelo.

El método de conjunto de usar Voting Classifier en muestras plegadas en K estratificadas brindó una precisión de > 64%

Durante la segunda parte del modelado para lograr el objetivo esperado, se necesito generar un predictor de la calidad del agua en esta parte nos decidimos por el modelo de GradientBoostingClassifier, y fue ligeramente mas tardado que el anterior, sin embargo, si logramos una mejor precisión de un 80% en la predicción necesaria.

4. CONCLUSIONES

Para cerrar la revisión de los datos encontrados y logrados, se puede decir que hemos logrado encontrar la información con aquellos datos que incidieran más directamente en lo bebible que un cuerpo de agua pueda ser esto es muy importante para la salud como se mencionó anteriormente gracias a esto podríamos finalmente lograr evitar muchas enfermedades graves en general de tipo infecciosas, donde la primera línea de batalla se habría vencido y de esta manera impediríamos que se desperdicien vidas y recursos en arreglar un problema aún mayor.

Sobre los modelos y el código, esta vez en comparación con otras oportunidades los modelos requirieron mas tiempo de procesamiento para lograr resultados satisfactorios es por eso que se procedieron a repetir los modelos que definimos tendrían una mayor precisión para este tipo de base de datos. Aunque finalmente los resultados fueron lo suficientemente aceptables.

REFERENCIAS

ASTM (1998) ASTM annual book of standards. Vol. 11.02. Philadelphia, PA, American Society for Testing and Materials.

kinetic phosphorimetry. Analytical Chemistry, 64:1413–1418.

D. K. Detterman. A challenge to watson. Intelligence, page 7778, 2011.

T.G. Evans. A program for the solution of a class of geometric-analogy intelligencetest questions. pages 271–353, 1968.

Peter D. Grunwald and Paul M. B. Vitányi. Algorithmic information theory. 2008.